Spatial filtering optimisation in motor imagery EEG-based BCI

Tetiana Aksenova, Alexandre Barachant, Stephane Bonnet

SPATIAL FILTERING OPTIMISATION IN MOTOR IMAGERY EEG-BASED BCI

Tetiana Aksenova, Alexandre Barachant, Stéphane Bonnet
CEA, LETI/DTBS/STD/LE2S
17 rue des Martyrs
38054 Grenoble cedex 09
tetiana.aksenova@cea.fr, alexandre.barachant@cea.fr, stephane.bonnet@cea.fr

ABSTRACT
Common spatial pattern (CSP) is becoming a standard way to combine linearly multi-channel EEG data in order to increase discrimination between two motor imagery tasks. We demonstrate in this article that the use of **robust estimates** allows improving the quality of CSP decomposition and CSP-based BCI. Furthermore, a scheme for electrode subset selection is proposed. It is shown that CSP with such subset of electrodes provides better results with the ones obtained with CSP over large multi-channel recordings.

KEY WORDS
Spatial filter – Robust statistics – Motor imagery – Brain-Computer Interface

1. Introduction

The movement-related Brain-Computer Interfaces (BCIs) aim at providing an alternative non-muscular communication path and control system for the individuals with severe motor disability to send the command to the external world using the measures of brain activity. Recently, several approaches and methods were developed to face the problem of brain movement-related signal decoding [1]. Non-invasive BCIs use mainly electroencephalographic activity (EEG) recorded from the scalp. In particular the power changes in various frequency bands are used to discriminate classes of EEG signal corresponding to the different types of motor activities.

It is well-recognized that the Common Spatial Pattern (CSP) algorithm is useful to increase the discriminative power of classifiers [2]. However, it was demonstrated in a recent study [3] that CSP is sensitive to outliers in the estimation of the intra-class covariance matrix. For instance, eye or tongue movements, muscle contractions are well-known sources of artefacts. Moreover the lack of concentration of the BCI operators may lead to non reliable motor task trials. This trial rejection in the learning stage is not addressed in [3] and should be accounted for robust CSP learning.

Finally, CSP is influenced by the spatial resolution of the acquisition system – number of electrodes and electrode location. Applying CSP processing to a large set of electrodes may also lead to overtraining by giving artificial weights to electrodes that do not convey information for the tasks under consideration.

In this paper, a new robust CSP (mcdcov-CSP) algorithm is proposed for intra-class centre and covariance matrix estimation. Better performance of the proposed algorithm is demonstrated in comparison with conventional cov-CSP. The use of subset of electrodes allows increasing the discrimination performance due to the elimination of non-informative electrodes.

2. Methods

2.1 CSP algorithm

CSP is a well-known approach in BCI systems to combine linearly multi-channel EEG data in order to increase discrimination between **two** motor imagery tasks [2]. The CSP algorithm finds the best projection matrix W such that the projected EEG signals have maximal (respectively minimal) variances for one class while the variance of the other class is minimized (respectively maximized). The variances of the projected signals are the features used during the subsequent classification stage.

The CSP algorithm can be formulated as a simultaneously diagonalization of the two intra-class spatial covariance matrices Σ_1, Σ_2 under an equality constraint.

$$\begin{cases} D = W^T \Sigma_1 W \\ I = W^T (\Sigma_1 + \Sigma_2) W \end{cases} \quad (1)$$

CSP algorithm is usually based on **sample covariance matrix estimation** and is therefore prone to errors due to the presence of possible outliers in the data.

The spatially filtered signals are obtained via the projection $\tilde{X} = W^T X$ where X denotes the EEG data recording, represented as a $E \times T$ matrix with E the number of electrodes and T the number of samples. Each column vector w_j, $j = 1...E$ of W, is called **a spatial filter** and is associated to the eigenvalue d_j, the j-th diagonal element of D. The most significant filters may be obtained by sorting the absolute distances $|d_j - 0.5|$ in decreasing order and keeping only the largest ones.

2.2. Robust CSP

A. Robust covariance matrix estimation

The starting point of a CSP analysis is the estimation of intra-class covariance matrices Σ_1 and Σ_2. The classes $i=1,2$ are represented by the set of EEG recording (trials) $\{X_j\}_{j\in S_i}$ where S_i is the set of trial indices corresponding to the i-th class.

First, covariance matrix is estimated for each trial and then intra-class mean covariance matrices are computed by averaging. First step is usually done using classical sample covariance estimates:

$$\hat{\Sigma} = \mathrm{cov}(X) = E\left\{(X-\bar{X})(X-\bar{X})^T\right\} \quad (2)$$

It is the maximum likelihood estimator when each column of X : x_i is an observation independently drawn from an *E*-variate normal distribution. For robust scatter matrix estimation, several approaches have been proposed. Minimum Covariance Determinant (MCD) is one of the most popular since it has a high breakdown point (α=0.5) which means that the algorithm successfully treats the trials corrupted by up to 50% of outliers [4]. Scatter matrix is estimated by the sample covariance matrix applied onto the subset of *h* observations which yields the lowest possible determinant. It also provides a robust estimate of the location $\hat{\mu}_{MCD}$. FAST-MCD algorithm allows avoiding a complete enumeration of all *h*-subsets out of *T* [5].

The MCD-based estimates of multivariate location and scatter allow defining a robust distance for each column observation x_i to the centre

$$RD(x_i) = \sqrt{(x_i-\hat{\mu}_{MCD})^T \hat{\Sigma}_{MCD}^{-1}(x_i-\hat{\mu}_{MCD})} \quad (3)$$

and rejecting outliers which are above a cut-off value $RD(x_i) \geq \sqrt{\chi^2_{E,0.975}}$. This value insures keeping realizations within a 97.5% robust confidence ellipse.

B. Robust intra-class covariance matrix

The next step of CSP is an estimate of interclass scatter matrices by averaging operation. In spite of robustness of MCD, it may occur that some trials – used during the averaging process - have a low specificity or they may be too much contaminated by artefacts. Rejecting these trials during the learning stage is also of primary importance.

To eliminate such irrelevant trials, MCD approach is again proposed. Let us note that in case of normality $x_i \sim N(0,\Sigma)$ the empirical covariance matrix $\hat{\Sigma}$ follow a Wishart distribution with (T-1) degree of freedom $\hat{\Sigma} \sim W_E(\Sigma, T-1)$ [6]. In addition, the marginal distribution of its diagonal elements s_{ii} is a χ^2 distribution with (T-1) degree of freedom $s_{ii}/\sigma_{ii} \sim \chi^2_{T-1}$ where σ_{ii} is the standard deviation of the i–th electrode [6]. That means that $\sqrt{2(s_{ii}/\sigma_{ii})} - \sqrt{2(T-1)-1}$ asymptotically follow to standard normal distribution $N(0,1)$. This approximation could be efficiently used for T >30 [7]. Due to asymptotical normality of $\sqrt{s_{ii}}$, we apply MCD algorithm to the transformed vector of the diagonal elements: $\sqrt{s_{ii}}$. Then, we use the robust distance (3) to detect abnormal trials and reject them during the averaging procedure.

C. Classification

After (robust) CSP, the variances of the projected signals are the features used during the classification stage achieved by Linear Discriminant Analysis (LDA). Two robust estimators of variances were tried in this study – a) the diagonal elements of covariance matrix of $\tilde{X}$ obtained with MCD algorithm and b) using Median Absolute Deviation [3]:

$$\hat{s}_i = \left[\frac{1}{0.6745} med\left(\left|\tilde{x}_i - med(\tilde{x}_i)\right|\right)\right]^2 \quad (5)$$

Robust estimators of variances will diminish the effect of outliers in the projected EEG signals. Moreover EEG recordings could show some differences to normality and thus the scatter matrix estimation proposed in previous subsection could be biased. The use of the same MCD algorithm to estimate variances will better correspond to the learning stage.

2.3. Electrodes subset selection

Finally, the use of a large set of electrodes may lead to overtraining by giving artificial weights to electrodes. Dimensionality reduction provides better performance of classifiers on the independent (test) dataset. For the electrodes subset selection, we applied simple correlation-based algorithm that includes sequentially the electrodes with highest correlation coefficient to the selected set. The procedure is stopped using “left corner” rule, if the coefficient of multiple correlation stops to increase essentially.

3. Results

3.1. Data description

To test our algorithm, several computational experiments were carried out. The datasets from the BCI competition III provided by Fraunhofer FIRST (Intelligent Data Anzalysis Group) and University Medicine Berlin (Neurophysics Group) [8] were used for testing. EEG data presents two classes which correspond to the right hand and the right foot motor imageries. Data is recorded with 118 electrodes with sampling rate 100Hz from 5 subjects (only subject *al*

and subject *av* are considered here) and for 280 trials. During the experiments the subject was given visual cues that indicate 3.5s time interval to perform motor imagery. In this study, we considered the band-pass filtered EEG signals in the [8-35] Hz band during the first 2.5 s (T=250) Furthermore, we only used two spatial filters.

To test the robustness of the algorithm, outliers were simulated as a mixture of distributions [9].

$$\xi_t \sim (1-\varepsilon)\delta_0 + \left(\frac{\varepsilon}{2}\right)N(\mu, \Sigma_{out}) + \left(\frac{\varepsilon}{2}\right)N(-\mu, \Sigma_{out}) \quad (6)$$

Here $N(\mu, \Sigma_{out})$ denotes the multivariate normal distribution, δ_0 is a point mass distribution and $\varepsilon \leq 1$ represents the part of the data corrupted by outliers. We set for Σ_{out} the diagonal matrix whose i-th diagonal element is the estimated variances $\hat{\sigma}_i^2$ for the i-th electrode. The mean amplitude vector of outliers (positive or negative) μ was also fixed according to the electrode standard deviation $\mu_i = \kappa \cdot \hat{\sigma}_i$ with κ a multiplicative factor. The outliers were added to the training trials as an additive noise, cf. Fig. 1.

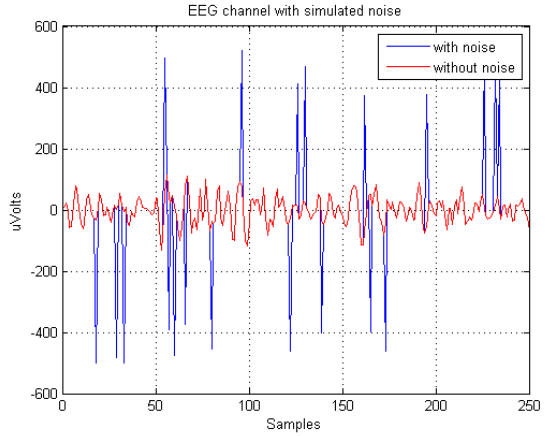

Figure 1. Example of EEG channel #10 with simulated noise: outlier amplitude is $\mu_{10} = 10\hat{\sigma}_{10}$, the part of corrupted observations is $\varepsilon = 0.1$.

A. *Electrode subset selection*

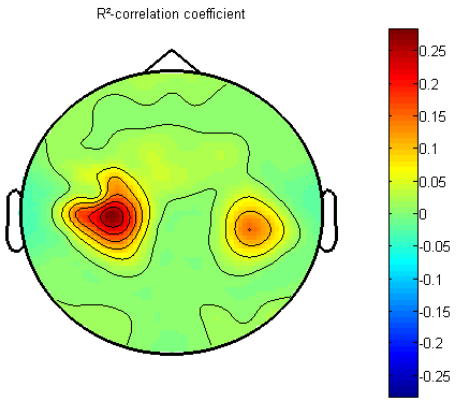

Figure 2. Spatial distribution of the R^2 correlation coefficient for patient *al*.

Fig. 2 demonstrates the cortex areas which are correlated to the motor imagery task. This simple criterion allows us to select the best subset of electrodes for a given mental task. We have then ordered the electrodes according to their R^2-value and evaluate the corresponding classification performance. As it is shown in Fig. 3, restricting our analysis to the first 21 electrodes gives comparable results to the whole electrode set. Indeed it may be the case that applying CSP processing to a large set of electrodes may lead to overtraining by giving artificial weights to electrodes that do not convey information for the tasks under consideration. An effect of overtraining is clearly observed in patient *av* for large number of electrodes.

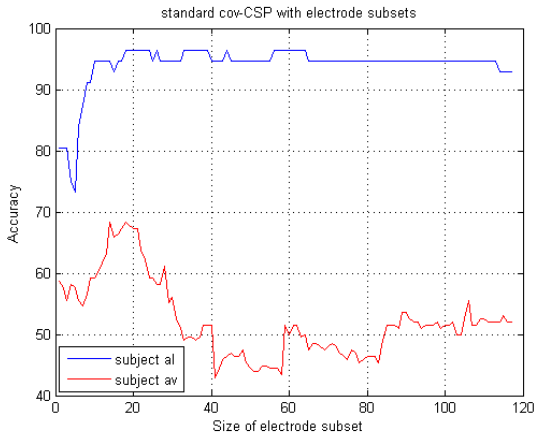

Figure 3. Classification accuracy using test dataset for 2 subjects (al: 20% test trial, av: 70% test trial)

B. Cov-CSP and robust (mcdcov-CSP) comparison.

The resistance of cov-CSP and mcdcov-CSP algorithms is studied depending on the outliers parameters.

a) Mean outlier amplitudes are fixed to $\mu_i = 3\hat{\sigma}_i$ while the percentage of corrupted observations increases from 0% to 25%. The results are demonstrated in Fig. 4 for subject *al*. We notice that mcdcov-CSP successfully resist to 25% of outliers while the accuracy degrades linearly using cov-CSP. Using MAD-estimation for the variance (mcdcov-mad-CSP method) yields also robust classification results.

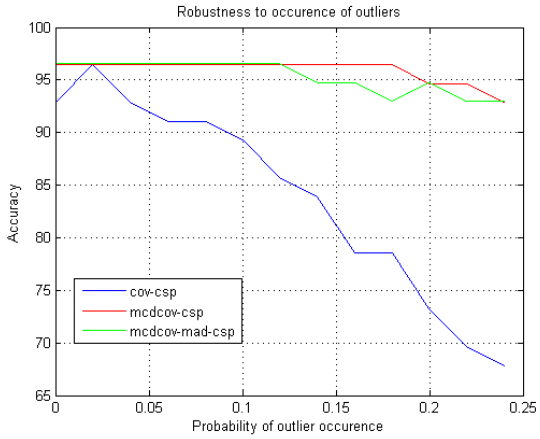

Figure 4. The percentage of correctly classified test trials (patient *al*), depending on the probability of outliers occurrence, κ=3.

It is also of interest to check the robustness of the spatial pattern in the presence/absence of noise (ε=0.1, κ=10). We observe in Fig. 5. that the most significant spatial pattern remains stable with the proposed approach. On the other hand, standard CSP with noise yields a perturbed spatial pattern (top-right figure).

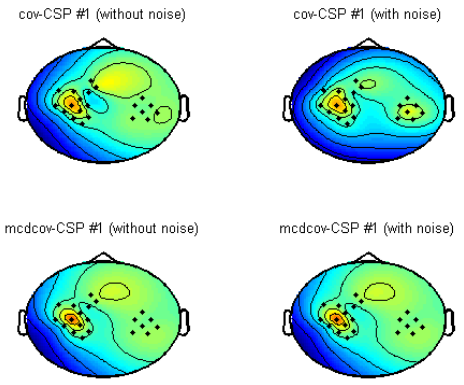

Figure 5. Most significant spatial pattern obtained with cov-CSP (top row: left without noise, right: with noise) and with mcdcov-CSP (top row: left without noise, right: with noise).

b) The percentage of the corrupted observation is fixed at 10% while the mean outlier amplitude coefficient factor κ varies from 0 to 100. The results, shown in Fig. 6, confirm that the mcdcov-CSP method and its variant are completely insensitive to high-amplitude outliers.

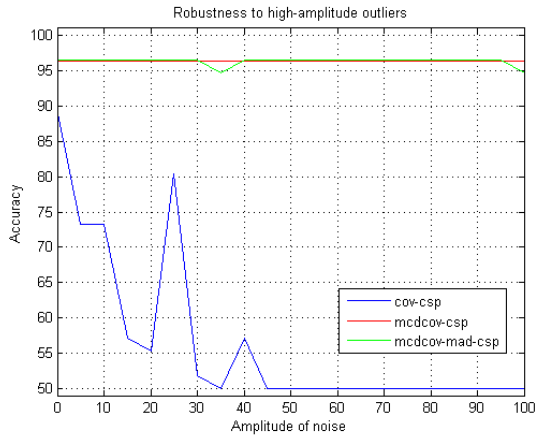

Figure 6. The percentage of correctly classified test trials (patient *al*), depending on the mean outlier amplitude and at a constant probability of occurrence.

3. Conclusion

The use of *robust estimates* allows improving the quality of CSP decomposition and CSP-based BCI. Robust mcdcov-CSP successfully resist to 25% of outliers while the standard cov-CSP degrade significantly. In addition, dimension reduction is important from a computational point of view since robust statistical methods are time-consuming. It is demonstrated that comparable results can be achieved with a well-selected subset of electrodes and it allows avoiding overtraining effects.

Acknowledgements

This work has been achieved within the CLINATEC project conducted in CEA/LETI/DTBS at Grenoble.

References

[1] M.A. Lebedev, and M.A.L. Nicolelis, Brain–machine interfaces: past, present and future, *TRENDS in Neurosciences 29*(9), 2006, 536-546.

[2] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, *IEEE Signal Proc Magazine*, *25*(1), 2008, 41-56.

[3] X. Yong, R.K. Ward, G.E. Birch, Robust common spatial patterns for EEG signal preprocessing, *Proc 30th annual intl IEEE EMBS Conf.*, Vancouver, Canada, 2008, 2087-2090.

[4] P.J. Rousseeuw, K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, *Technometrics 41*, 1999, 212 -223.

[5] S. Verboven, M. Hubert, LIBRA: a MATLAB Library for Robust Analysis, 2004, (http://www.wis.kuleuven.ac.be/stat/robust.html)

[6] K. V. Mardia, J. T Kent, and J. M. Bibby, *Multivariate Analysis* (Academic Press, Duluth, London, 1979).

[7] G.A. Korn, T.M. Korn, *Mathematical handbook*, (McGraw Hill Book Company, 1968)

[8] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans. Biomed. Eng.*, *51*(6), 2004, 993-1002.

[9] P. Huber, *Robust Statistics*, (John Wiley & Sons Inc., 2003)